Final Retail Model Replication
July 24, 2007

This final guidance document presents all variable calculation and filter information, along with the final regression format, coefficients, and lookup table.

1. Notes on Variable Calculation

The final model requires a variety of calculations on the CBECS data in order to apply filters and put all of the variables in the right format. The following notes detail these calculations.

Energy Calculations

- Recodes: Observations that report "no use" for a given fuel type are not asked for a quantity. Therefore, for these observations it is appropriate to recode the system missing values to zero.
- Source Energy: The following factors are applied to convert values in CBECS into Source Energy
  - Electricity = 3.34
  - Natural Gas = 1.047
  - Fuel Oil = 1.01
  - Propane = 1.01
  - District Heat = 1.45 (if STUSED8=1)
  - District Heat = 1.35 (if only Hot Water is used, STUSED8=2)
- Propane: The amount of propane (in gallons) must be estimated from the category selected. This will involve a "maximum" estimation for the purpose of applying filters and an "actual" estimation for use in the rating models. For the maximum, the high end of the propane category is used. For the actual, the middle of the category is used. This is summarized in the table below.

Table 1
Propane Estimation Guidelines

PRAMTC8	Maximum Estimation (filter)	Actual Estimation (model)
1: less than 100	99	50
2: 100 to 499	499	300
3: 500 to 999	999	750
4,7,8,9: 1000 or higher, or unknown	Not Included	Not Included

In order to convert the amount of propane in gallons or pounds (indicated by PRUNIT8) into a Site energy value in kBtu, the following conversions were applied:

- 1 Gallon of Propane = 92 kBtu
- 1 Gallon of Propane = 4.5 Pounds

General Attribute Calculations
The following variables are necessary for the calculation of the model:

- Source Energy: the sum of each form of energy in a building multiplied by the source multipliers presented in Section 1
- Source Intensity: SourceEUI = Source Energy/Sqft8
- Natural Log of Square foot: LNSqFt = LN(SqFt8)
- Worker Density[1] = WkrDen = 1000*NWker8/SqFt8
- Computer Density: PCDen: 1000*PCNUM8/SqFt8
- Heating Degree Days x Percent Heated: HDDxPH = HDD658*HeatP8/100
- Cooling Degree Days x Percent Cooled: CDDxPC = CDD658*CoolP8/100
- Cash Register Density: RgstrDen = 1000*RGSTRN8/SqFt8
- Walk in Refrigeration Density: WalkinDen = 1000*RfgWIN8/SqFt8
- Open and Closed Refrigeration Density:
  CommDen = 1000*(RFGOPN8 + RFGCLN8)/SqFt8

2. Filters

Filters are established for a couple of reasons: to account for technical limitations of the CBECS data, to maintain baseline operational characteristics deemed appropriate for the program, and to prevent outliers from driving the regression result. The following table summarizes the filters for the Retail model, alongside the rationale for each filter and the number of observations that remain after each filter. Once all of the filters have been applied, there are 182 observations remaining.

[1] All operating characteristic density values are computed on a per 1000 square foot basis

Table 2 Summary of Retail Model Filters		
Condition for *Including* an Observation in the Analysis	**Rationale**	**Number Remaining**
PBAPLUS8=42	All Retail Stores – Starting Set	291
Must have a weight (value for ADJWT8)	CBECS Limitation – Cannot incorporate un-weighted observations into analysis	291
Must have square foot <=1,000,000	CBECS Limitation – Actual values above 10^6 are not reported	291
If propane is used, the amount category (PRAMTC8) must equal 1, 2, or 3	CBECS Limitation – Cannot estimate propane use if it is "greater than 1000" or unknown	281
If propane is used, the maximum estimated propane amount must be 10% or less of the total source energy[2]	CBECS Limitation – Estimation of propane cannot introduce too much error into the energy use value	274
Must not use chilled water	CBECS Limitation – Quantities of chilled water are not collected	272
Must operate for at least 30 hours per week	EPA Requirement – Baseline condition for being a full time retail building	263
Must operate for at least 10 months per year	EPA Requirement – Baseline condition for being a full time retail building	249
Retail activity must characterize greater than 50% of the floor space[3]	EPA Requirement – In order to be considered part of the retail peer group, more than 50% of the building must be defined by retail activity	241
Must have square foot >= 5,000	Analytical Limitation – Analysis could not model behavior for these smaller buildings	182

3. Variable Centering

For the purpose of the model, all of the variables are centered. Centering creates a situation where the intercept is equal to the average source energy intensity and the coefficients are used to adjust a building based on its deviation from the average value of each operating characteristic.

- Centered Variable = Observation Value – Weighted Mean for that Variable
- Centered LnSqFt = LnSqFt – Weighted Average LnSqFt
- Note that the weighted averages are computed across the filtered data set

Final centering values are presented in the following table.

[2] Here, the high end of the propane usage amount category (PRAMT8) is used to determine a maximum propane consumption and estimate what percent this would be of the total source energy. If that percent exceeds 10% of the total, then the observation is removed from the analysis.

[3] Generally, this is determined by a couple of screens. If the variable ONEACT8=1, this indicates that one activity is in 75% or more of the building. If the variable ONEACT8=2, then the observation can specify up to 3 activities (ACT18, ACT28, ACT38). One of these activities must be retail (PBAX8=15), and must account for more than 50% of the floor area.

Table 3 Final Centering Values		
Variable	**Full Name**	**Center Value (Weighted Mean)**
LNSqFt	Natural Log of Square foot	9.370678
WkHrs	Weekly Operating Hours	63.74061
WkrDen	Number of Workers per 1000 square feet	0.627964
PCDen	Number of Computers per 1000 square feet	0.314887
HDDxPH	Heating Degree Days x Percent Heated	3810.862
CDDxPC	Cooling Degree Days x Percent Cooled	972.0578
RgstrDen	Number of Registers per 1000 square feet	0.190495
WalkinDen	Number of Walk-in Refrigerators per 1000 square feet	0.003779
RfgCommDen	Number of Open and Closed Refrigerators per 100 square feet	0.045045
Note: *Computed over the filtered data set (n=182 observations), see Section 2* *Values are weighted by ADJWT8*		

4. Regression

The final regression is a weighted ordinary least squares regression across the filtered data set of 182 observations. The dependent variable is SourceEUI. Each independent variable is centered as described in Section 3. The final model is estimated as follows.

Table 4 Model Summary Output - SPSS				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.842(a)	.710	.695	2219.17386

a Predictors: (Constant), C_RfgCommDen, C_LNSqFt, C_HDDxPH, C_WalkinDen, C_WkrDen, C_CDDxPC, C_RgstrDen, C_PCDen, C_Wkhrs

Table 5 Model ANOVA Output - SPSS					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	2071690775.110	9	230187863.901	46.741	.000(a)
Residual	847054007.722	172	4924732.603		
Total	2918744782.832	181			

a Predictors: (Constant), C_RfgCommDen, C_LNSqFt, C_HDDxPH, C_WalkinDen, C_WkrDen, C_CDDxPC, C_RgstrDen, C_PCDen, C_Wkhrs

b Dependent Variable: SrcEUI

c Weighted Least Squares Regression - Weighted by ADJWT8 ADJWT8

Table 6 Model Coefficient Estimates Output - SPSS				
	Unstandardized Coefficients	**Standard Error**	**t**	**Sig.**
(Constant)	153.099	5.685	26.931	.000
C_LNSqFt	20.187	9.315	2.167	.032
C_Wkhrs	1.373	.421	3.263	.001
C_WkrDen	61.755	15.536	3.975	.000
C_PCDen	70.601	20.804	3.394	.001
C_HDDxPH	.01127	.003	4.274	.000
C_CDDxPC	.01255	.007	1.725	.086
C_RgstrDen	249.058	33.785	7.372	.000
C_WalkinDen	720.212	379.622	1.897	.059
C_RfgCommDen	81.903	44.342	1.847	.066

a Dependent Variable: SrcEUI

b Weighted Least Squares Regression - Weighted by ADJWT8 ADJWT8

5. Lookup Table

The model is used to generate a scoring lookup table according to the following steps:

- The model is used to generate a predicted Source EUI for each observation.
- An energy efficiency ratio is calculated for each observation as follows:

  Actual Source EUI/Predicted Source EUI
- The weighted cumulative distribution of the energy efficiency ratios is fitted to a gamma distribution, where the parameters alpha and beta are estimated in order to minimize the sum of square differences between the actual and the gamma value for the cumulative distribution of each observation[4].
- For the retail data the final gamma distribution is computed to have the following values:
  - Alpha = 4.2595
  - Beta = 0.2397
  - Sum of square error = 0.0740
- The validity of this fit can be verified graphically, as shown below.
- The final gamma shape and scale parameters (alpha and beta, respectively) are used to calculate the efficiency ratio at each percentile (1 to 100) along the curve. These ratios are the lookup table, as shown below.

[4] This fit can be achieved using MS Excel's Solver function. The actual cumulative distribution (weighted) is computed for each observation. Then, arbitrary alpha and beta values are assigned and used to compute a gamma distribution value for each observation. The solver can be set to minimize the sum of the differences between the actual and gamma distribution values across all observations, by changing both alpha and beta. The resulting alpha and beta are the final shape and scale values, respectively.

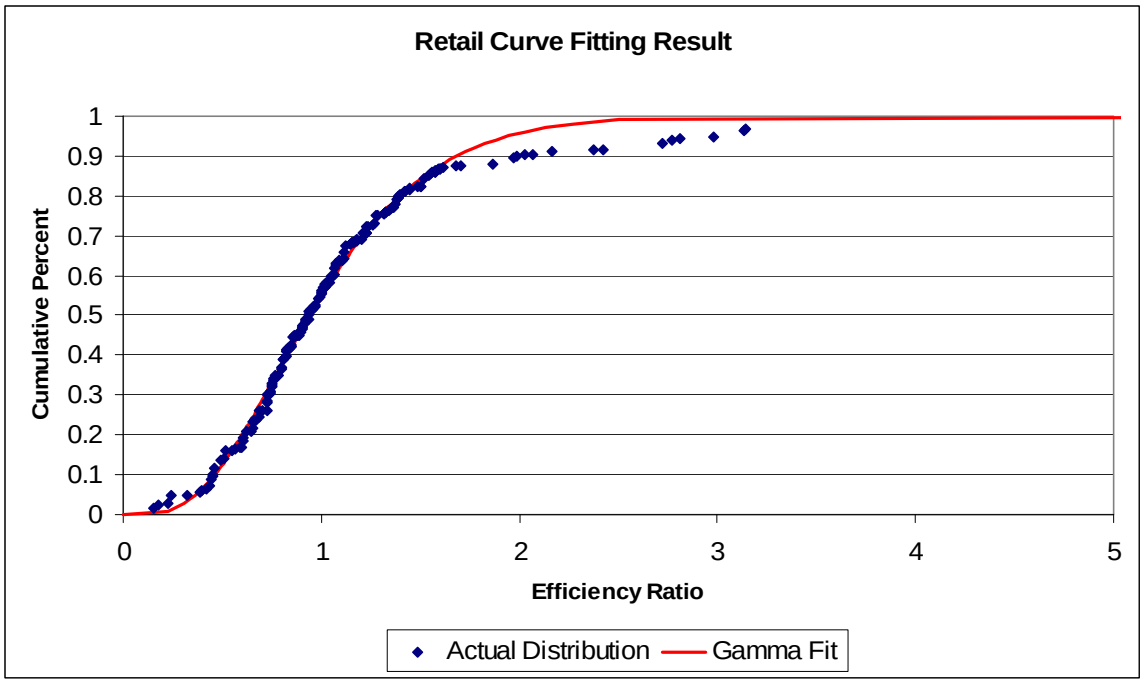
Retail Curve Fitting Result
Cumulative Percent
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
Efficiency Ratio
Actual Distribution
Gamma Fit

Table 7 Lookup Table for Retail Rating					
Rating	**Cumulative Percent**	**Ratio**	**Rating**	**Cumulative Percent**	**Ratio**
100	0	<0.2243	50	0.5	0.9543
99	0.01	0.2743	49	0.51	0.9662
98	0.02	0.3100	48	0.52	0.9783
97	0.03	0.3390	47	0.53	0.9906
96	0.04	0.3640	46	0.54	1.0029
95	0.05	0.3864	45	0.55	1.0154
94	0.06	0.4067	44	0.56	1.0281
93	0.07	0.4256	43	0.57	1.0409
92	0.08	0.4434	42	0.58	1.0539
91	0.09	0.4602	41	0.59	1.0671
90	0.1	0.4762	40	0.6	1.0804
89	0.11	0.4915	39	0.61	1.0940
88	0.12	0.5063	38	0.62	1.1078
87	0.13	0.5206	37	0.63	1.1219
86	0.14	0.5345	36	0.64	1.1362
85	0.15	0.5480	35	0.65	1.1507
84	0.16	0.5612	34	0.66	1.1656
83	0.17	0.5742	33	0.67	1.1808
82	0.18	0.5868	32	0.68	1.1963
81	0.19	0.5993	31	0.69	1.2121
80	0.2	0.6116	30	0.7	1.2284
79	0.21	0.6237	29	0.71	1.2450
78	0.22	0.6356	28	0.72	1.2621
77	0.23	0.6474	27	0.73	1.2797
76	0.24	0.6591	26	0.74	1.2978
75	0.25	0.6707	25	0.75	1.3164
74	0.26	0.6822	24	0.76	1.3357
73	0.27	0.6936	23	0.77	1.3556
72	0.28	0.7050	22	0.78	1.3762
71	0.29	0.7162	21	0.79	1.3977
70	0.3	0.7275	20	0.8	1.4200
69	0.31	0.7387	19	0.81	1.4433
68	0.32	0.7499	18	0.82	1.4677
67	0.33	0.7610	17	0.83	1.4933
66	0.34	0.7721	16	0.84	1.5202
65	0.35	0.7833	15	0.85	1.5488
64	0.36	0.7944	14	0.86	1.5791
63	0.37	0.8055	13	0.87	1.6115
62	0.38	0.8167	12	0.88	1.6463
61	0.39	0.8279	11	0.89	1.6841
60	0.4	0.8391	10	0.9	1.7252
59	0.41	0.8503	9	0.91	1.7707
58	0.42	0.8616	8	0.92	1.8216
57	0.43	0.8729	7	0.93	1.8795
56	0.44	0.8843	6	0.94	1.9469
55	0.45	0.8958	5	0.95	2.0280
54	0.46	0.9073	4	0.96	2.1307
53	0.47	0.9189	3	0.97	2.2721
52	0.48	0.9306	2	0.98	2.5066
51	0.49	0.9424	1	0.99	>2.5066